Jethva Parthiv

+91 90819 08259 | jethva.parthiv007@gmail.com | linkedin.com/in/jethva-parthiv | github.com/Jethva-Parthiv | Portfolio

SUMMARY

AI Engineer with hands-on experience building and deploying agentic AI applications, RAG systems, and fine-tuned NLP model. Developed OpsPilot (Autonomous RevOps Agent) and BankIntent AI, a fine-tuned DistilBERT intent classifier model achieving 90.83% test accuracy. Currently optimizing NoteFlow (Voice-to-Notion Agent) for lower latency and improving harness reliability. Independently owns the full lifecycle — design, debugging, evaluation, and deployment — with projects live and publicly demoable. Also built other AI projects available on GitHub. Seeking an AI Engineer internship to contribute to real-world AI systems from day one.

TECHNICAL SKILLS

Languages & Concepts: Python, SQL, OOPs, DSA, Asynchronous Programming

Agent Harness: LLMs API Integration, LangGraph, LangChain, Tool Calling, Chatbots, Embeddings, Multi-Agent Systems, Prompt Engineering, Observability, LLM Evals, MCP, STT-TTS

Core Concepts: Agents, RAG, Machine Learning, Deep Learning, NLP

Frameworks & Tools: FastAPI, Flask, Streamlit, React (basic), REST APIs, Pydantic

Libraries: Pytorch, Pandas, NumPy, NLTK, Scikit-Learn, HuggingFace Transformers, FastMCP, Matplotlib

Databases: PostgreSQL, MySQL, Qdrant, FAISS, ChromaDB, Pinecone, Redis

MLOps & LLMOps: LangSmith, Weights & Biases, Docker, Streamlit Cloud, HuggingFace Spaces, AWS (Console Basics) , Render, Vercel

Tools: Git, GitHub, Postman, Google Colab, VS Code, Antigravity, n8n

PROJECTS

OpsPilot — Autonomous RevOps AI Agent Platform | *Demo* | *GitHub*

- **Tech Stack:** Python, FastAPI, LangGraph, LangChain, Google Gemini, React (Vite + Tailwind CSS), ChromaDB, SQLAlchemy, LangSmith, Slack SDK, Notion API, Salesforce API, SSE.
- Architected a ReAct loop agent in LangGraph with Gemini function calling, dynamically selecting across 4 tools (Notion RAG, Salesforce lookup, CRM task creation, entity extraction).
- Built an event-driven platform with Slack Socket Mode, FastAPI backend, a React dashboard with live SSE updates, and LangSmith observability for full pipeline tracing
- Designed a Human-in-the-Loop approval workflow letting operators review and edit AI-drafted responses before Slack publication.
- Validated performance with a 3-tier eval suite: 94% triage accuracy, 90% RAG top-3 hit rate, 4.2/5.0 reply quality (LLM-as-judge).

BankIntent AI — Customer Support Ticket Classifier | *Live* | *GitHub*

- **Tech Stack:** DistilBERT, PyTorch, HuggingFace, W&B, Gradio
- Fine-tuned DistilBERT on the Banking77 dataset (10,003 samples, 77 intent classes) using a custom PyTorch training loop written from scratch, achieving 90.83% test accuracy and 0.908 weighted F1.
- Tracked experiments end-to-end in Weights & Biases, including live loss curves, per-epoch F1 scores, and interactive 77×77 confusion matrix charts.
- Built an interactive Gradio web application with dynamic Top-K intent ranking, latency telemetry, and automated department triage routing deployed to Hugging Face Spaces.

NoteFlow — Autonomous Voice-to-Notion Capture Agent | *GitHub*

- **Tech Stack:** Python, FastAPI, LangGraph, Google Gemini, Notion MCP, Groq Whisper, edge-tts, LangSmith
- Engineered a Voice-Driven Notion Agent: Developed an autonomous assistant using FastAPI, LangGraph, and Google Gemini connected to Notion MCP, transcribing voice via Groq Whisper and delivering spoken confirmations via edge-tts.
- Cut capture latency by over 70% by pairing an in-memory workspace cache with prompt-tuned single-turn creation for new notes, while maintaining full multi-turn ReAct loops for exploratory multi-step page updates.
- Built LangGraph duplicate-tool interception and recursion circuit breakers to stop infinite search loops, implemented English-locked, zero-temperature STT to reduce hallucinations, and used LangSmith to trace agent trajectories, tool calls, and token usage.

RAGVerse AI — End-to-End RAG Framework | *GitHub*

- **Tech Stack:** Python, LangChain, Google Gemini, ChromaDB, PostgreSQL, RAGAS, LangSmith.
- Designed a modular RAG system using LangChain, enabling multi-turn conversations with persistent memory.
- Built a ChromaDB retrieval pipeline with hybrid retrieval, reranking, and configurable chunking, backed by PostgreSQL checkpointing for resumable sessions.
- Integrated Gemini for response generation and built a RAGAS-based eval pipeline measuring faithfulness, relevancy, and context precision.
- Structured the system as a state-based workflow with LangSmith observability, Supporting multi-agent use cases.

EDUCATION

Master of Computer Applications (MCA) — **Pursuing** — 2025 – 2027 (Expected)
Gujarat Vidyapith, Ahmedabad, Gujarat

Bachelor of Computer Applications (BCA) — 2022 – 2025 , **CGPA : 7.58**
Hemchandracharya North Gujarat University (HNGU), Patan, Gujarat

ACHIEVEMENTS & CERTIFICATIONS | View Certificates

Programming Competitions & Hackathons

- **1st Place** — Python Programming Competition | Tech-Fest | Sardar Patel University
- **Winner** — Code Seige (Python) | Tech Kaushalya | Gujarat University
- **Runner-up** — Logic Forge (Python) | Tech Kaushalya | Gujarat University
- **1st Place** — Code Combat (Java) | IGNITE CMPICA | CHARUSAT
- **Participant** — n8n Community Hackathon | FCA, CMPICA, CHARUSAT
- **Participant** — SSIP Regional Hackathon & New India Vibrant Hackathon | Ganpat University

Certifications

- **AWS Training & Certification** — Introduction to Generative AI: Art of the Possible